Belle Preprint 2002-28
KEK Preprint 2002-83
TIT-HPE-02-01

Measurement of the Electroweak Penguin Process $B \to X_s \ell^+ \ell^-$

J. Kaneko,[47] K. Abe,[9] K. Abe,[44] T. Abe,[45] I. Adachi,[9] Byoung Sup Ahn,[16] H. Aihara,[46] M. Akatsu,[23] Y. Asano,[51] T. Aso,[50] V. Aulchenko,[2] T. Aushev,[13] A. M. Bakich,[41] Y. Ban,[34] E. Banas,[28] W. Bartel,[6] A. Bay,[19] P. K. Behera,[52] A. Bondar,[2] A. Bozek,[28] M. Bračko,[21,14] J. Brodzicka,[28] T. E. Browder,[8] B. C. K. Casey,[8] P. Chang,[27] Y. Chao,[27] K.-F. Chen,[27] B. G. Cheon,[40] R. Chistov,[13] Y. Choi,[40] Y. K. Choi,[40] M. Danilov,[13] L. Y. Dong,[11] S. Eidelman,[2] V. Eiges,[13] Y. Enari,[23] C. W. Everton,[22] F. Fang,[8] H. Fujii,[9] C. Fukunaga,[48] N. Gabyshev,[9] A. Garmash,[2,9] T. Gershon,[9] R. Guo,[25] J. Haba,[9] F. Handa,[45] T. Hara,[32] Y. Harada,[30] N. C. Hastings,[22] H. Hayashii,[24] M. Hazumi,[9] E. M. Heenan,[22] I. Higuchi,[45] T. Higuchi,[46] L. Hinz,[19] T. Hojo,[32] Y. Hoshi,[44] W.-S. Hou,[27] H.-C. Huang,[27] T. Igaki,[23] Y. Igarashi,[9] T. Iijima,[23] K. Inami,[23] A. Ishikawa,[23] H. Ishino,[47] R. Itoh,[9] H. Iwasaki,[9] Y. Iwasaki,[9] H. K. Jang,[39] J. H. Kang,[55] J. S. Kang,[16] N. Katayama,[9] H. Kawai,[3] Y. Kawakami,[23] N. Kawamura,[1] T. Kawasaki,[30] H. Kichimi,[9] D. W. Kim,[40] Heejong Kim,[55] H. J. Kim,[55] H. O. Kim,[40] Hyunwoo Kim,[16] S. K. Kim,[39] K. Kinoshita,[5] S. Kobayashi,[37] S. Korpar,[21,14] P. Križan,[20,14] P. Krokovny,[2] R. Kulasiri,[5] S. Kumar,[33] Y.-J. Kwon,[55] J. S. Lange,[7,36] G. Leder,[12] S. H. Lee,[39] J. Li,[38] A. Limosani,[22] R.-S. Lu,[27] J. MacNaughton,[12] G. Majumder,[42] F. Mandl,[12] D. Marlow,[35] S. Matsumoto,[4] T. Matsumoto,[48] W. Mitaroff,[12] K. Miyabayashi,[24] Y. Miyabayashi,[23] H. Miyake,[32] G. R. Moloney,[22] T. Mori,[4] T. Nagamine,[45] Y. Nagasaka,[10] T. Nakadaira,[46] E. Nakano,[31] M. Nakao,[9] J. W. Nam,[40] K. Neichi,[44] S. Nishida,[17] O. Nitoh,[49] S. Noguchi,[24] T. Nozaki,[9] S. Ogawa,[43] T. Ohshima,[23] T. Okabe,[23] S. Okuno,[15] S. L. Olsen,[8] Y. Onuki,[30] W. Ostrowicz,[28] H. Ozaki,[9] P. Pakhlov,[13] H. Palka,[28] C. W. Park,[16] H. Park,[18] K. S. Park,[40] J.-P. Perroud,[19] M. Peters,[8] L. E. Piilonen,[53] N. Root,[2] K. Rybicki,[28] H. Sagawa,[9] Y. Sakai,[9] H. Sakamoto,[17] M. Satapathy,[52] A. Satpathy,[9,5] O. Schneider,[19] S. Schrenk,[5] C. Schwanda,[9,12] S. Semenov,[13] K. Senyo,[23] R. Seuster,[8] H. Shibuya,[43] B. Shwartz,[2] V. Sidorov,[2] J. B. Singh,[33] N. Soni,[33] S. Stanič,[51,*] K. Sumisawa,[9] T. Sumiyoshi,[48] K. Suzuki,[9] S. Suzuki,[54] S. Y. Suzuki,[9] S. K. Swain,[8] H. Tajima,[46] T. Takahashi,[31] F. Takasaki,[9] K. Tamai,[9] N. Tamura,[30] J. Tanaka,[46] M. Tanaka,[9] G. N. Taylor,[22] Y. Teramoto,[31] S. Tokuda,[23] M. Tomoto,[9] T. Tomura,[46] K. Trabelsi,[8] W. Trischuk,[35,†] T. Tsuboyama,[9] T. Tsukamoto,[9] S. Uehara,[9] K. Ueno,[27] S. Uno,[9] Y. Ushiroda,[9] G. Varner,[8] K. E. Varvell,[41] C. C. Wang,[27] C. H. Wang,[26] J. G. Wang,[53] M.-Z. Wang,[27] Y. Watanabe,[47] E. Won,[16] B. D. Yabsley,[53] Y. Yamada,[9] A. Yamaguchi,[45] Y. Yamashita,[29] M. Yamauchi,[9] H. Yanai,[30] J. Yashima,[9] M. Yokoyama,[46] Y. Yuan,[11] Y. Yusa,[45] C. C. Zhang,[11] J. Zhang,[51] Z. P. Zhang,[38] Y. Zheng,[8] V. Zhilich,[2] and D. Žontar[51]

(The Belle Collaboration)
[1]Aomori University, Aomori
[2]Budker Institute of Nuclear Physics, Novosibirsk
[3]Chiba University, Chiba
[4]Chuo University, Tokyo
[5]University of Cincinnati, Cincinnati OH
[6]Deutsches Elektronen–Synchrotron, Hamburg
[7]University of Frankfurt, Frankfurt
[8]University of Hawaii, Honolulu HI
[9]High Energy Accelerator Research Organization (KEK), Tsukuba
[10]Hiroshima Institute of Technology, Hiroshima
[11]Institute of High Energy Physics,
Chinese Academy of Sciences, Beijing
[12]Institute of High Energy Physics, Vienna
[13]Institute for Theoretical and Experimental Physics, Moscow
[14]J. Stefan Institute, Ljubljana
[15]Kanagawa University, Yokohama
[16]Korea University, Seoul
[17]Kyoto University, Kyoto
[18]Kyungpook National University, Taegu
[19]Institut de Physique des Hautes Énergies, Université de Lausanne, Lausanne
[20]University of Ljubljana, Ljubljana
[21]University of Maribor, Maribor
[22]University of Melbourne, Victoria
[23]Nagoya University, Nagoya
[24]Nara Women's University, Nara
[25]National Kaohsiung Normal University, Kaohsiung
[26]National Lien-Ho Institute of Technology, Miao Li
[27]National Taiwan University, Taipei
[28]H. Niewodniczanski Institute of Nuclear Physics, Krakow
[29]Nihon Dental College, Niigata
[30]Niigata University, Niigata
[31]Osaka City University, Osaka
[32]Osaka University, Osaka
[33]Panjab University, Chandigarh
[34]Peking University, Beijing
[35]Princeton University, Princeton NJ
[36]RIKEN BNL Research Center, Brookhaven NY
[37]Saga University, Saga
[38]University of Science and Technology of China, Hefei
[39]Seoul National University, Seoul
[40]Sungkyunkwan University, Suwon
[41]University of Sydney, Sydney NSW
[42]Tata Institute of Fundamental Research, Bombay
[43]Toho University, Funabashi
[44]Tohoku Gakuin University, Tagajo
[45]Tohoku University, Sendai

[46]*University of Tokyo, Tokyo*
[47]*Tokyo Institute of Technology, Tokyo*
[48]*Tokyo Metropolitan University, Tokyo*
[49]*Tokyo University of Agriculture and Technology, Tokyo*
[50]*Toyama National College of Maritime Technology, Toyama*
[51]*University of Tsukuba, Tsukuba*
[52]*Utkal University, Bhubaneswer*
[53]*Virginia Polytechnic Institute and State University, Blacksburg VA*
[54]*Yokkaichi University, Yokkaichi*
[55]*Yonsei University, Seoul*
(Dated: August 18, 2002)

Abstract

We report the first measurement of the branching fraction for the inclusive decay $B \to X_s \ell^+ \ell^-$, where ℓ is either an electron or a muon and X_s is a hadronic recoil system that contains an s-quark. We analyzed a data sample of 65.4×10^6 B meson pairs collected with the Belle detector at the KEKB e^+e^- asymmetric-energy collider. We find $\mathcal{B}(B \to X_s \ell^+ \ell^-) = (6.1 \pm 1.4(\mathrm{stat})\,^{+1.4}_{-1.1}(\mathrm{syst})) \times 10^{-6}$ for dilepton masses greater than 0.2 GeV/c^2.

PACS numbers: 13.20.He, 14.65.Fy, 14.40.Nd

The first observation of the flavor-changing-neutral-current (FCNC) weak decay process $B \to K\ell^+\ell^-$, recently reported by Belle [1], opens a new window for searches for physics beyond the Standard Model (SM) [2]. There are no SM first-order weak decay processes that can produce such decays; the dominant SM processes are second-order b-quark to s-quark processes with a virtual t-quark in a loop (electroweak penguin) or a box diagram. Non-SM processes could produce sizable modifications to the decay amplitude due to contributions from virtual non-SM particles (such as charged Higgs or SUSY particles) in the loop.

The decay amplitude is often described by an effective Hamiltonian in which non-SM contributions can modify the Wilson coefficients C_7, C_9 and C_{10}. The magnitude of allowable modifications of C_7 is constrained by the measured rate for inclusive $B \to X_s\gamma$ decays [3, 4, 5]. The Belle measurement of the exclusive process $B \to K\ell^+\ell^-$ has been used to determine the best limits on non-SM contributions to C_9 and C_{10} [6]. The usefulness of exclusive measurements is limited by the large theoretical uncertainties associated with the s-quark to K meson hadronization process. Since these uncertainties are not as severe for inclusive processes, measurements of the branching fraction and the dilepton and hadronic mass spectra for $B \to X_s\ell^+\ell^-$ will provide more stringent and less model-dependent probes for new physics. This process has not been measured; CLEO has reported a 90% confidence upper limit of $\mathcal{B}(B \to X_s\ell^+\ell^-) < 4.2 \times 10^{-5}$ [7].

In this Letter, we present the results of a measurement of the branching fraction for the inclusive decay $B \to X_s\ell^+\ell^-$, where B is either B^0 or B^+, ℓ is either an electron or a muon, and X_s is a hadronic recoil system that contains an s-quark. Here and throughout the paper, the inclusion of charge conjugated modes is implied. We use a data sample collected at the $\Upsilon(4S)$ resonance with the Belle detector at the KEKB e^+e^- asymmetric-energy collider (3.5 GeV on 8 GeV) [8]. This sample contains $(65.4 \pm 0.5) \times 10^6$ B meson pairs, corresponding to an integrated luminosity of 60 fb^{-1}.

The Belle detector is a large-solid-angle magnetic spectrometer that consists of a three-layer silicon vertex detector (SVD), a 50-layer central drift chamber (CDC), an array of aerogel threshold Čerenkov counters (ACC), time-of-flight scintillation counters (TOF), and an electromagnetic calorimeter comprised of CsI(Tl) crystals (ECL) located inside a super-conducting solenoid coil that provides a 1.5 T magnetic field. An iron flux-return located outside of the coil is instrumented with resistive plate counters to identify muons (KLM). The detector is described in detail elsewhere [9].

We reconstruct charged particle trajectories with the CDC and SVD. Electron identification is based on the position and shower shape of the cluster in the ECL, ratio of the cluster energy to the track momentum (E/p), specific energy-loss measurement (dE/dx) with the CDC and the response from the ACC. Muon identification is based on the hit positions and the depth of penetration into the ECL and KLM. We require the electron's (muon's) laboratory momentum to be greater than 0.5 GeV/c (1 GeV/c). We find an electron (muon) selection efficiency of 92.5% (91.3%) with a (0.2 ± 0.06)% ((1.4 ± 0.04)%) pion to electron (muon) mis-identification probability. Charged kaon candidates are selected by using information from the ACC, TOF and dE/dx in the CDC. The kaon selection efficiency is 90% with a pion to kaon mis-identification probability of 6%. The remaining charged tracks are assumed to be pions. We select $K^0_S \to \pi^+\pi^-$ candidates with invariant mass within 15 MeV/c^2 of the K^0 mass. We impose additional K^0_S selection criteria based on the distance and the direction of the K^0_S vertex and the impact parameters of daughter tracks. We require the charged tracks other than those used in the K^0_S reconstruction to have impact parameters with respect to the nominal interaction point of less than 0.5 cm in the radial

direction and 3 cm along the beam direction.

We reconstruct photons from ECL energy clusters that have no associated charged tracks. The shower shape is required to be consistent with an electromagnetic cluster and the energy to be greater than 50 MeV. We reconstruct $\pi^0 \to \gamma\gamma$ candidates from photon pairs with invariant mass within 10 MeV/c^2 of the nominal π^0 mass.

The X_s system is reconstructed in 18 different combinations of either a K^+ or K_S^0 combined with 0 to 4 pions, of which up to one π^0 is allowed. We combine the X_s with two oppositely charged leptons to form a B candidate. We identify the $B \to X_s\ell^+\ell^-$ signal with the beam-energy constrained mass, $M_{\rm bc} = \sqrt{(E_{\rm beam}^{\rm CM}/c^2)^2 - (p_B^{\rm CM}/c)^2}$, where $E_{\rm beam}^{\rm CM}$ and $p_B^{\rm CM}$ are the beam energy and the B candidate momentum calculated in the center-of-mass (CM) system, respectively. We find the average $M_{\rm bc}$ resolution is $\sigma_{\rm bc} = 2.8$ MeV/c^2. The energy difference $\Delta E = E_B^{\rm CM} - E_{\rm beam}^{\rm CM}$, where $E_B^{\rm CM}$ is the CM energy of the B candidate, is combined with other variables to provide background suppression.

To optimize the selection criteria and to determine their signal efficiencies, we use the following signal model. The SM calculation of Ref. [6] is used for the dilepton mass ($M_{\ell^+\ell^-}$) spectrum. We require $M_{\ell^+\ell^-}$ to be greater than 0.2 GeV/c^2 to remove the virtual photon contribution, $b \to s\gamma^* \to se^+e^-$ [10]. We model the recoil mass (M_{X_s}) spectrum using a non-resonant Fermi motion model [11] and the JETSET hadronization program [12]. For $M_{X_s} <$ 1.1 GeV/c^2, however, the spectrum is modeled by the sum of the exclusive $B \to K^{(*)}\ell^+\ell^-$ components as predicted by the SM [13]. Using this model we find that our reconstructed final states account for 81% of the total X_s states. Figures 1(a) and (b) show the $M_{\ell^+\ell^-}$ and M_{X_s} spectra from this model.

There are two background sources that can peak in $M_{\rm bc}$ and ΔE. The first is hadronic B decays into one kaon plus multiple pions ($B \to X_s\pi^+\pi^-$) from abundant decays such as $B \to D^{(*)}n\pi$ ($n \geq 1$). We estimate this background contribution by reconstructing $B \to X_s\pi^+\pi^-$ events without the lepton identification criteria and multiplying by the measured momentum dependent pion mis-identification probability. We find contributions of 2.6 ± 0.2 events to the $B \to X_s\mu^+\mu^-$ signal and 0.1 ± 0.05 events to $B \to X_se^+e^-$. We refer to this as the fake background, and subtract it from the signal yield. The second is from $B \to J/\psi X_s$ and $B \to \psi' X_s$ where J/ψ and ψ' decay into dileptons. These decay modes have the same final states and, in principle, interfere with the signal. For this study, these charmonium decays are explicitly vetoed. The veto windows are -0.6 to $+0.2$ GeV/c^2 (-0.35 to $+0.2$ GeV/c^2) around the J/ψ mass for the e^+e^- ($\mu^+\mu^-$) channel, and -0.3 to $+0.15$ GeV/c^2 around the ψ' mass for both channels. We find the background from this source is negligibly small using a $B \to J/\psi X_s$ Monte Carlo (MC) sample in which the $M_{\ell^+\ell^-}$ spectrum reproduces that of data.

The largest background sources are random combinations of dileptons with a kaon and pions that originate from continuum $q\bar{q}$ ($q = u, d, s, c$) production or from semileptonic B decays. We reject 83% of the $q\bar{q}$ background with a signal efficiency of 90% by using a Fisher discriminant [14] ($\mathcal{F}_{\rm cont}$) based on a modified set of Fox-Wolfram moments [15] that differentiate the event topology. In the semileptonic B decay background, both B mesons decay into leptons or two leptons are produced from the $b \to c \to s, d$ decay chain. We combine the missing mass ($M_{\rm miss}$) and the total visible energy ($E_{\rm vis}$) [16] into another Fisher discriminant ($\mathcal{F}_{\rm sl}$) to reject 85% of events with two neutrinos with a signal efficiency of 91%.

We further reduce the backgrounds using ΔE and the cosine of the B flight direction ($\cos\theta_B$) with respect to the e^- beam direction in the CM frame. First, we select events with $|\Delta E| < 40$ MeV ($\sim 3\sigma_{\Delta E}$). We then calculate likelihoods $\mathcal{L}_{S,B} = p_{S,B}^{\Delta E} \times p_{S,B}^{\rm cosB}$, where

$p_{S,B}^{\Delta E}$ and $p_{S,B}^{\cos B}$ are the uncorrelated probability density functions (PDF) for ΔE and $\cos\theta_B$ for the signal (S) and the background (B), respectively, and form a likelihood ratio $\mathcal{LR} = \mathcal{L}_S/(\mathcal{L}_S + \mathcal{L}_B)$. We model $p_S^{\Delta E}$ with a Gaussian primarily determined from a signal MC sample and calibrated using $B \to J/\psi X_s$ data; $p_B^{\Delta E}$ is modeled with a linear function with a slope determined from a MC sample that contains $b \to c$ decays and $q\bar{q}$ events. The signal follows a $1 - \cos^2\theta_B$ distribution, while the background distributes uniformly in $\cos\theta_B$. By requiring $\mathcal{LR} > 0.8$, we retain 75% of the signal while removing 80% of backgrounds.

For events with multiple candidates that pass the selection criteria, we choose the combination with the largest value of $\mathcal{LR}$. We find that the correct candidate is reconstructed in 73% of the events. We reject candidates with X_s invariant mass greater than 2.1 GeV/c^2. This condition removes a large fraction of combinatorial background while retaining (93±5)% of the signal.

Signal MC samples are used to determine reconstruction efficiencies of (3.7 ±0.4 ±0.5)%, where the first error is systematic and the second error is due to the model uncertainty. Here, an equal production rate is assumed for $B^0\overline{B}^0$ and B^+B^-. Systematic errors include uncertainties from the tracking efficiency (2.0% per track), K_S^0 reconstruction efficiency (8.7% per K_S^0), π^0 reconstruction efficiency (6.8% per π^0), electron identification (1.8% per electron), muon identification (2.2% per muon), K^+ identification (2.5% per K^+), π^+ identification (0.8% per π^+), and the background suppression criteria (3.0% for $\mathcal{F}_{\rm cont}$, $\mathcal{F}_{\rm sl}$ and $\mathcal{LR}$, combined, estimated using $B \to J/\psi X_s$ data).

The model uncertainty includes the following sources. The largest error is due to uncertainties in the SM branching fraction of the exclusive modes (11%). The uncertainty in the M_{X_s} spectrum (4%) is due to the Fermi momentum (p_F) and spectator quark mass (m_q) parameters. We take $p_F = 0.4$ GeV/c and $m_q = 0$ GeV/c^2, and vary the parameters over a range allowed by the measured heavy quark effective theory (HQET) parameters λ_1 and $\overline{\Lambda}$ [4, 17]. The uncertainty in the fraction of the unmeasured modes (2.1%) and the uncertainty due to the fractions of π^0 and K_S^0 contained in X_s (4.6%) are estimated by comparing the inclusive hadron (π^0, K_S^0, η, ϕ, etc.) production rates between continuum data and JETSET.

We determine the signal yield from an unbinned maximum likelihood fit to the $M_{\rm bc}$ distribution as shown in Fig. 2. We model the signal with a Gaussian function and the background with a threshold function [18]. The width of the signal Gaussian is obtained from $B \to J/\psi X_s$ data. The background shape is obtained from the background MC sample. We verify that the threshold function obtained from MC reproduces the shape of data for $B \to X_s e\mu$ combinations (Fig. 2(d)). The fit results are given in Table I; we find $60.1 \pm 13.9({\rm stat})\,^{+8.6}_{-5.4}({\rm syst})$ events for the combined $B \to X_s\ell^+\ell^-$ modes (Fig. 2(c)) with a statistical significance of 5.4σ. Here, the systematic error is obtained from the maximum deviation when the background and signal shapes are varied by one standard deviation of the statistical error in their shape parameters. The significance is defined as $\sqrt{-2\ln(\mathcal{L}_0/\mathcal{L}_{\rm max})}$, where $\mathcal{L}_{\rm max}$ is the maximum likelihood and $\mathcal{L}_0$ is the maximum likelihood when the signal yield is constrained to be zero. The fake background and the uncertainty in the background shape are included in the significance calculation. We calculate the branching fraction for $M_{\ell^+\ell^-} > 0.2$ GeV/c^2 to be

$$\mathcal{B}(B \to X_s\ell^+\ell^-) = (6.1 \pm 1.4({\rm stat})\,^{+1.4}_{-1.1}({\rm syst})) \times 10^{-6},$$

where the systematic errors in the yield, efficiency, the number of B meson pairs and the model errors are added in quadrature to give the total systematic error. This can be com-

pared to the SM expectation $\mathcal{B}(B \to X_s\ell^+\ell^-) = (4.2 \pm 0.7) \times 10^{-6}$ [10]. Table I summarizes the branching fractions for $B \to X_s e^+e^-$ and $B \to X_s\mu^+\mu^-$ separately, together with the number of candidates, signal yields, fake background estimations, efficiencies and the statistical significances.

The dilepton mass spectrum is measured by dividing the data into $M_{\ell^+\ell^-}$ bins. For each bin, the signal yield is extracted from a fit to the $M_{\rm bc}$ distribution and the fake background contribution is subtracted. The result is shown in Fig. 1(c). Similarly, the recoil mass spectrum is obtained by dividing the data into M_{X_s} bins and extracting the signal yield for each bin. The result is shown in Fig. 1(d). With the current statistics, the $M_{\ell^+\ell^-}$ and M_{X_s} spectra are in agreement with SM expectations. Branching fractions for each bin are given in Table II.

In summary, we present the first measurement of the inclusive branching fraction for the electroweak penguin decay $B \to X_s\ell^+\ell^-$. The results are in agreement with the SM expectations and can be used to constrain extensions of the SM.

We wish to thank the KEKB accelerator group for the excellent operation of the KEKB accelerator. We thank A. Ali, E. Lunghi and G. Hiller for providing us many helpful suggestions and calculations. We acknowledge support from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the Japan Society for the Promotion of Science; the Australian Research Council and the Australian Department of Industry, Science and Resources; the National Science Foundation of China under contract No. 10175071; the Department of Science and Technology of India; the BK21 program of the Ministry of Education of Korea and the CHEP SRC program of the Korea Science and Engineering Foundation; the Polish State Committee for Scientific Research under contract No. 2P03B 17017; the Ministry of Science and Technology of the Russian Federation; the Ministry of Education, Science and Sport of the Republic of Slovenia; the National Science Council and the Ministry of Education of Taiwan; and the U.S. Department of Energy.

---

* on leave from Nova Gorica Polytechnic, Nova Gorica

† on leave from University of Toronto, Toronto ON

[1] K. Abe *et al.* (Belle Collaboration), Phys. Rev. Lett. **88**, 021801 (2002); the BaBar collaboration also reports preliminary results in B. Aubert *et al.* (BaBar Collaboration), hep-ex/0207082, unpublished.

[2] For example, E. Lunghi, A. Masiero, I. Scimemi and L. Silverstrini, Nucl. Phys. **B568**, 120 (2000); J. L. Hewett and J. D. Wells, Phys. Rev. **D55**, 5549 (1997); T. Goto, Y. Okada, Y. Shimizu and M. Tanaka, Phys. Rev. **D55**, 4273 (1997); G. Burdman, Phys. Rev. **D52**, 6400 (1995); N. G. Deshpande, K. Panose and J. Trampetić, Phys. Lett. **B308**, 322 (1993); W. S. Hou, R. S. Willey and A. Soni, Phys. Rev. Lett. **58**, 1608 (1987).

[3] K. Abe *et al.* (Belle Collaboration), Phys. Lett. **B511**, 151 (2001).

[4] S. Chen *et al.* (CLEO Collaboration), Phys. Rev. Lett. **87**, 251807 (2001).

[5] R. Barate *et al.* (ALEPH Collaboration), Phys. Lett. **B429**, 169 (1998).

[6] A. Ali, E. Lunghi, C. Greub and G. Hiller, Phys. Rev. **D66**, 034002 (2002); A. Ali, P. Ball, L. T. Handoko and G. Hiller, Phys. Rev. **D61**, 074024 (2000). These papers use results that include partial next-to-next-to-leading logarithmic QCD corrections from C. Bobeth, M. Misiak and J. Urban, Nucl. Phys. **B574**, 291 (2000) and H. H. Asatrian, H. M. Asatrian,

C. Greub and M. Walker, Phys. Lett. **B507**, 162 (2001).
[7] S. Glenn *et al.* (CLEO Collaboration), Phys. Rev. Lett. **80**, 2289 (1998).
[8] E. Kikutani ed., KEK Preprint 2001-157 (2001), to appear in Nucl. Instr. and Meth. A.
[9] A. Abashian *et al.* (Belle Collaboration), Nucl. Instr. and Meth. **A479**, 117 (2002).
[10] The original authors of Ref. [6] calculate $\mathcal{B}(B \to X_s e^+e^-) \sim \mathcal{B}(B \to X_s \mu^+\mu^-) \sim (4.2 \pm 0.7) \times 10^{-6}$ for $M_{\ell^+\ell^-} > 0.2$ GeV/c^2.
[11] A. Ali, G. Hiller, L. T. Handoko and T. Morozumi, Phys. Rev. **D55**, 4105 (1997).
[12] T. Sjöstrand, "PYTHIA 5.6 and JETSET 7.3: Physics and manual", CERN-TH-6488-92.
[13] We use $\mathcal{B}(B \to K\ell^+\ell^-) = (3.5 \pm 1.2) \times 10^{-7}$ and $\mathcal{B}(B \to K^*\ell^+\ell^-) = (1.2 \pm 0.4) \times 10^{-6}$ from Ref. [6].
[14] R. A. Fisher, Annals Eugen. **7**, 179 (1936).
[15] The Fox-Wolfram moments were introduced in G. C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978). The modified moments used by Belle are described in [3].
[16] We define $M_{\rm miss} = \sqrt{(2E_{\rm beam}^{\rm CM} - \sum E_i^{\rm CM})^2/c^4 - |\sum \vec{p}_i^{\,\rm CM}/c|^2}$ and $E_{\rm vis} = \sum E_i^{\rm CM}$, where the sum is over all the charged particles with a pion mass hypothesis and all photons.
[17] D. Cronin-Hennessy *et al.* (CLEO Collaboration), Phys. Rev. Lett. **87**, 251808 (2001).
[18] H. Albrecht *et al.* (ARGUS Collaboration), Phys. Lett. **B229**, 304 (1989).

TABLE I: Fit results for the number of candidates, signal yields, fake backgrounds, reconstruction efficiencies, statistical significances and branching fractions ($\mathcal{B}$). Candidates are counted in a $\pm 3\sigma_{\rm bc}$ window in $M_{\rm bc}$.

mode	candidates	signal yield	fake background	efficiency (%)	significance	$\mathcal{B}(\times 10^{-6})$
$B \to X_s e^+e^-$	96	$25.5 \pm 11.2\,^{+4.8}_{-3.8}$	0.1 ± 0.05	$3.9 \pm 0.4 \pm 0.5$	3.4	$5.0 \pm 2.3\,^{+1.3}_{-1.1}$
$B \to X_s \mu^+\mu^-$	92	$37.3 \pm 9.7\,^{+7.2}_{-3.8}$	2.6 ± 0.2	$3.6 \pm 0.4 \pm 0.5$	4.7	$7.9 \pm 2.1\,^{+2.1}_{-1.5}$
$B \to X_s \ell^+\ell^-$	188	$60.1 \pm 13.9\,^{+8.6}_{-5.4}$	2.7 ± 0.2	$3.7 \pm 0.4 \pm 0.5$	5.4	$6.1 \pm 1.4\,^{+1.4}_{-1.1}$

TABLE II: Branching fractions ($\mathcal{B}$) for each bin of $M_{\ell^+\ell^-}$ and M_{X_s}. The first and second errors are statistical and systematic, respectively.

$M_{\ell^+\ell^-}$ (GeV/c^2)	$\mathcal{B}(\times 10^{-7})$	M_{X_s} (GeV/c^2)	$\mathcal{B}(\times 10^{-7})$
[0.2, 1.0]	$9.3 \pm 7.4\,^{+1.7}_{-1.3}$	[0.4, 0.6]	$4.9 \pm 1.4\,^{+1.0}_{-0.8}$
[1.0, 2.0]	$11.2 \pm 6.6\,^{+2.7}_{-2.3}$	[0.6, 0.8]	$-0.6 \pm 1.0\,^{+0.8}_{-0.5}$
[2.0, $M_{J/\psi}$]	$18.7 \pm 7.6\,^{+4.5}_{-3.7}$	[0.8, 1.0]	$5.2 \pm 3.4\,^{+1.2}_{-1.0}$
[$M_{J/\psi}$, $M_{\psi'}$]	$10.3 \pm 3.7\,^{+2.5}_{-2.1}$	[1.0, 1.6]	$26.8 \pm 9.5\,^{+5.7}_{-4.6}$
[$M_{\psi'}$, 5.0]	$11.1 \pm 5.4\,^{+2.8}_{-2.2}$	[1.6, 2.1]	$26.8 \pm 13.8\,^{+6.0}_{-4.4}$

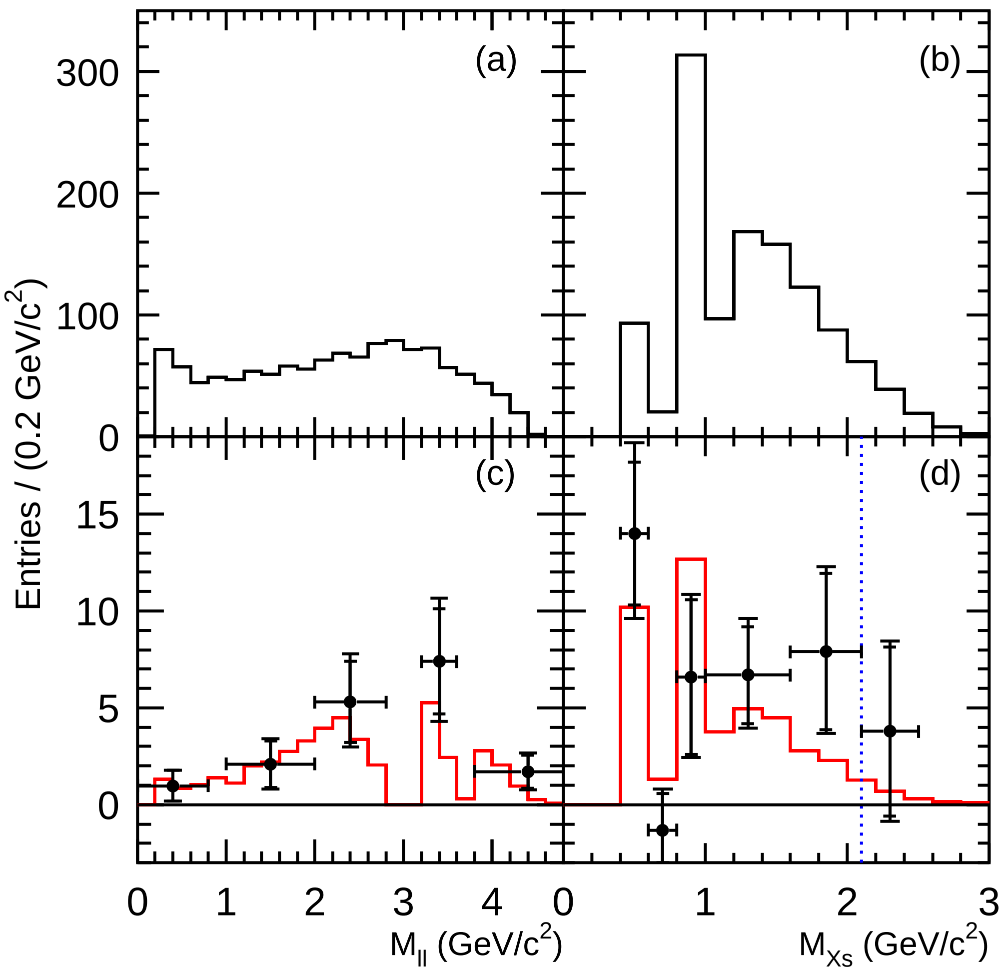

FIG. 1: SM expectations for the (a) dilepton and (b) recoil mass spectra; the observed (c) dilepton and (d) recoil mass spectra (circles). Inner and outer error bars indicate the statistical and total errors, respectively. The histograms in (c), (d) show the SM expectations after all the selections are applied; histograms are normalized to the expected branching fractions. The gaps in (c) are due to the J/ψ, ψ' vetoes. The dotted line in (d) indicates the $M_{X_s} < 2.1$ GeV/c^2 requirement.

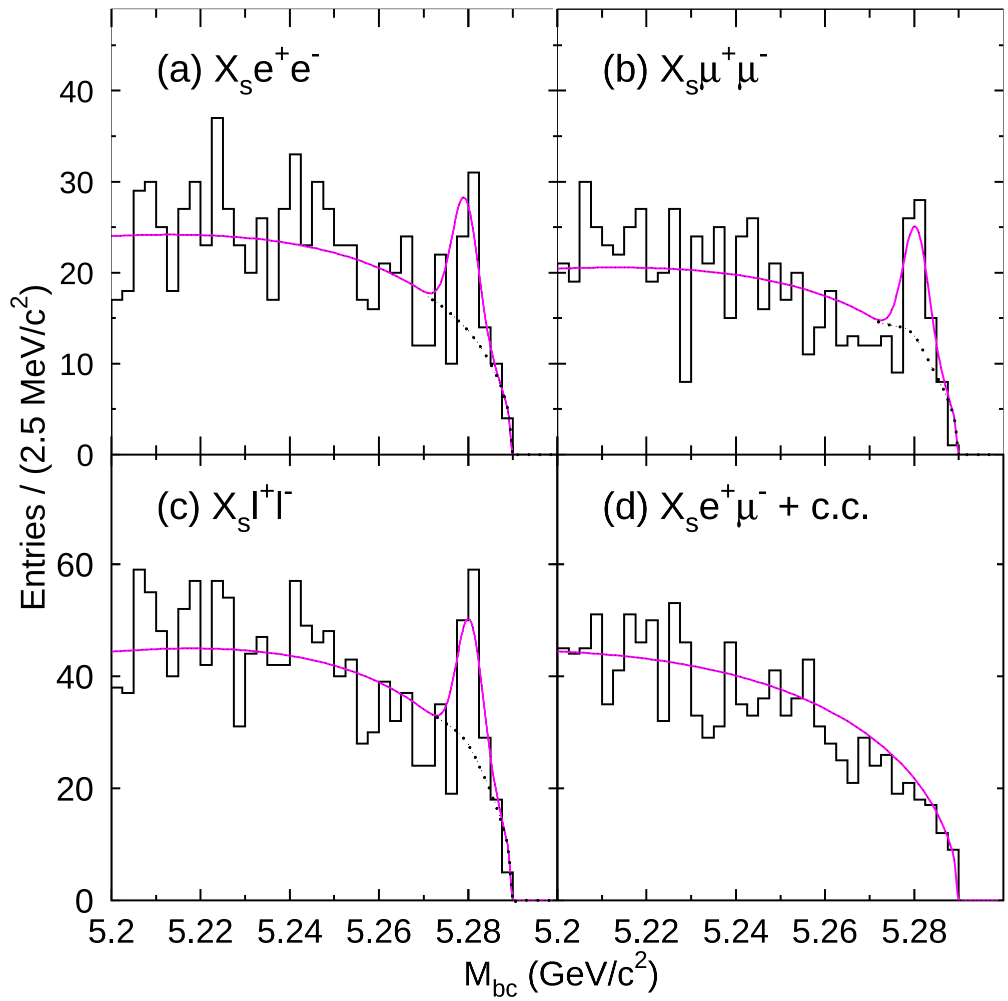

FIG. 2: Beam-energy constrained mass distributions for (a) $B \to X_s e^+e^-$, (b) $B \to X_s \mu^+\mu^-$, (c) $B \to X_s \ell^+\ell^-$ and (d) $B \to X_s e^\pm \mu^\mp$. The solid lines indicate the fit results and the dotted lines show the sum of the background components.